National Institute of Standards and Technology (NIST): Building Bridges in Biomedical Informatics and Computational Biology

I. HIV Structural Database

Brief description of your initiative and its goals:
HIV database provides significant advances in the annotation and dissemination of structural and ligand data for AIDS research with particular emphasis to industrial (drug discovery) interests. This Standard Reference Database (HIV SRD 102) for AIDS research incorporates state-of-the-art annotation tools and techniques developed at NIST specifically to manage structural information and its interoperability. It is useful in developing new AIDS inhibitors by facilitating the online comparison of the existing hundreds of AIDS inhibitors on the basis of their ability to attack specific locations in the active site of the AIDS enzyme (HIV protease). HIV SRD 102 includes not only data from the widely used RCSB Protein Data Bank developed in collaboration with NIST, but also previously unpublished data from industrial and other laboratories. The HIV database was originally developed in collaboration with the National Cancer Institute and was expanded recently to include two-dimensional structural data (inhibitors for which no 3-D structural data does not exist) for 1167 potent HIV protease inhibitor compounds supplied by the National Institute of Allergy and Infectious Diseases (NIAID).

Contact Information: Talapady N Bhat, bhat@nist.gov (301 -975 5448), Anne Plant, anne.plant@nist.gov (301 -975 3124)
Biochemical Sciences Division, National Institute of Standards and Technology

Web site: http://xpdb.nist.gov/hivsdb/hivsdb.html

Brief Description of Informatics and Computational Components and their goals:
HIV structural database holds, annotates and distributes 3-D structural data (solved using X-ray and NMR methods) received both from the widely used RCSB Protein Data Bank and previously unpublished data from industrial and other laboratories. Further, the database includes structures of over 1000 potent inhibitors of HIV protease obtained primarily from published literature and provided to NIST by NIAID. Work to include additional 1000 inhibitors is in progress. To our knowledge HIV SRD 102 is the only public data resource for AIDS structures that integrates 3-D structural data with structural data of other potent inhibitors for which 3-D structures are not available. For this purpose, several novel techniques to cross-index and organize chemical data have been developed. One of the methods is called Chem-BLAST (Chemical Block Layered Alignment of Substructure Technique) that interactively compares inhibitors using their fragments. Chem-BLAST is a dictionary based method to organize drugs using concepts derived from commonly asked questions by Medicinal Chemists and Modelers. These concepts allow the development of superior search engines on the Web to enable better on-line experience. Using Chem-BLAST a Web user is able to query the entire collection of

inhibitors using visual images of their components or fragments (http://xpdb.nist.gov/hivsdb/advanced_query_files/slide0002.htm) a feature unique to HIV SRD 102. This way, during a query, Chem-BLAST circumvents the problem of specifying structures using their complicated and hard to reproduce chemical IUPAC names. The method Chem-BLAST also lays the foundation for the development of Semantic Web for Chemical Structures as illustrated in our Chemical Taxonomies use case at the W3C Web site (http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase).

Brief Description of Resources and Tools available for sharing:
HIV structural database distributes structural data for over 1000 potent inhibitors. These data are available for onsite examination, comparison or downloads. The novel method, Chem-BLAST used for this work has been recently published (Chemical compound navigator: a web-based Chem-BLAST, chemical taxonomy-based search engine for browsing compounds.
Prasanna MD, Vondrasek J, Wlodawer A, Rodriguez H, Bhat TN. Proteins. 2006 Jun 1;63(4):907-17). Techniques and tools used in this work are available for inspection, discussion and sharing.

Brief description of integrative efforts:
Currently efforts are underway to integrate Chem-BLAST with the efforts by the World Wide Web Consortium (W3C) to develop Semantic Web concepts, tools, ontology, and standards for chemical structures. Our goal is to develop standards, tool, and techniques for establishing chemical taxonomies and ontology with special emphasis to drug discovery. We are also heavily involved in integrating HIV SRD 102 with other AIDS related databases such as the one at NIAID. Integration of these two databases by direct cross-indexing is planned for the near future.

Opportunities for collaboration or synergy with the NCBCs:
Opportunities for collaboration with NCBC exists in several areas such as chemical genomics, chemical and structural databases, methodology of curating, classifying, comparing and querying drugs and drug targets. Opportunities exist to develop collaborations both in (a) developing tools, techniques, and data standards, (b) developing healthcare related databases including those that involve images.

II. Semantic Web for Chemical Genomics:

Brief description of your initiative and its goals:
The Semantic Web is a vision for an intelligent Web of the future and it is often suggested as the solution to the growing problem in managing and distributing the millions of data points generated by modern science such as rational structure-based drug-design, nanotechnology and chemical genomics. However, the progress towards this vision has been very limited for several reasons. Two of the major reasons for this slow progress are the sheer complexity of the problem in preparing and presenting data to a semantics-aware search engine and the lack of clear examples that illustrate a successful implementation. A third major reason for the slow progress

is the anticipated large quantum technological jump to be made by database providers to build new data handlers and associated search engines consistent with the Semantic Web vision. We have been focusing on these issues with particular emphasis to Chemical Genomics and Structure-based drug design for AIDS. Recently, we have posted a use case of chemical taxonomies for Semantic Web in the W3C Website.

Contact Information: Talapady N Bhat, bhat@nist.gov (301 -975 5448), Anne Plant, anne.plant@nist.gov (301 -975 3124)
Biochemical Sciences Division, National Institute of Standards and Technology

Web site: http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase
http://xpdb.nist.gov/hivsdb/hivsdb.html

Brief Description of Informatics and Computational Components and their goals:
Annotating and classifying chemical compounds using the semantics popularly accepted by researchers involved in drug design is crucial for the effective use of biological and structural data for technological growth and drug development. Drug design is a collective and coordinated effort involving several disciplines such as structural biology and medicinal chemistry. The semantics used by these disciplines may differ from one another. Several semantic representations of the views commonly used by these disciplines are analysed and used in developing a chemical taxonomies for Semantic web use case.
http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase#head-b17ab780bff7395fdf242be61c6ce50a584b796e

Brief Description of Resources and Tools available for sharing:
Techniques and tools used in developing chemical taxonomies for the Semantic Web use case are available for discussion and sharing. Some of these concepts have been recently published (Chemical compound navigator: a web-based Chem-BLAST, chemical taxonomy-based search engine for browsing compounds. Prasanna MD, Vondrasek J, Wlodawer A, Rodriguez H, Bhat TN. Proteins. 2006 Jun 1;63(4):907-17).

Brief description of integrative efforts:
Efforts are underway to integrate our efforts with the ongoing efforts on Semantic Web Development by the World Wide Web Consortium (W3C). This effort is done in collaboration with John Barkley from the Information Technology Laboratory (ITL) of NIST and Eric Neumann - Director of the Clinical Semantics Group, W3C. Integrated query on the structural data of AIDS inhibitors are planned between those distributed by NIST and NIAID. This integrative effort is expected to result in the extension of our Semantic Web efforts to the compounds held in the NIAID databases.

Opportunities for collaboration or synergy with the NCBCs:

Opportunities for collaboration with NCBC for developing the Semantic Web Technology exists in several healthcare related areas such as drug development, chemical genomics, chemical and structural databases, and bio-imaging.

III. Evolutionary Informatics: software and standards to support comparative analysis.

Brief description of your initiative and its goals:
Most problems in computational genome analysis— strategies for annotating genomes, identifying potential drug targets, analyzing SNPs, and so on— are problems in comparative biology, i.e., making inferences from comparisons of evolved things. The proper theoretical basis for comparative analysis is evolutionary theory, which provides a framework to convert the kinds of questions that arise in comparative analysis into well posed questions about evolutionary transitions in the states of characters over time, dependent on a phylogenetic tree. Yet, most genome analysis does not take advantage of this powerful framework. The main problem is the lack of informatics support. Our project aims to facilitate the broader use of evolutionary methods by developing standards and software that will facilitate flexible, convenient and reliable storage, management, visualization and analysis of data using trees and models of character change.

Contact Information: Arlin Stoltzfus (arlin.stoltzfus@nist.gov; 240 314 6208)
Biochemical Sciences Division, National Institute of Standards and Technology

Web site: http://www.molevol.org/nexplorer

Brief Description of Informatics and Computational Components and their goals:
Current efforts depend on NEXPL, an applications programming interface (API) for the NEXUS file format, developed in Perl. NEXUS is a legacy format designed originally to support comparative analysis. Though NEXUS is still a powerful format, one goal of our project is to develop a new and improved standard. NEXPL provides much of the functionality behind the Nexplorer server, which provides a graphical user interface to manipulate and develop views of sequence family data sets including sequences and phylogenies. Developing robust automated methods for generating sequence family data sets (with a rich set of information about a sequence family) is also part of the project.

Brief Description of Resources and Tools available for sharing:
Nexplorer is available for public use and provides access to thousands of pre-computed sequence family data sets, as described by Gopalan, Qiu, Chen and Stoltzfus (2006, "Nexplorer: phylogeny-based exploration of sequence family data", Bioinformatics, 22(1):120-121).

Brief description of integrative efforts:

Currently an effort is underway to integrate a NEXUS-like object called CDAT (character data and trees) with BioPerl, the extensive library of Perl functions used extensively in genome projects around the world. This will provide the glue to connect genomics efforts with the standards and tools needed to facilitate evolutionary analysis. In the next stage of the project we will work with developers of other major software projects, working through NESCENT (the National Evolutionary Synthesis Center) to improve current implementations of NEXUS and to develop interfaces to other software commonly used in comparative analysis.

Opportunities for collaboration or synergy with the NCBCs:
Opportunities for collaboration with the NCBCs exist in several areas, including developing an appropriate ontology for data used in evolutionary comparative analysis, and improving the scalability of methods for storage and visualization.

IV. Enzyme Thermodynamics Database:

Brief description of your initiative and its goals:
Thermodynamic data on enzyme-catalyzed reactions play an important role in the prediction of the extent of reaction and the position of equilibrium for any process in which these reactions occur. The importance of understanding the thermodynamics of these biochemical reactions was emphasized by Krebs and Kornberg in their monograph "A Survey of the Energy Transformations in Living Matter". Their monograph also contains a useful appendix on Gibbs energy data of biological interest and a table on the thermodynamics of enzyme-catalyzed reactions. However, the amount of data available at that time was extremely limited. Reviews on various aspects of this subject have subsequently appeared. Each of these reviews, however, has been limited in the extent of coverage given to this area and no comprehensive review exists. Thermodynamic information is also needed in biotechnology when one needs to optimize product yields and to calculate the energy requirements of a given reaction.

It is the aim of this project to provide a compilation of data on the thermodynamics of enzyme-catalyzed reactions. The data presented herein is limited to direct equilibrium and calorimetric measurements performed on these reactions under in vitro conditions. This is the principal thermodynamic information that is needed to determine the position of equilibrium of a given reaction.

Contact information: Robert Goldberg robert.goldberg@nist.gov **,** Talapady N Bhat, Bhat@nist.gov

Web site: http://xpdb.nist.gov/enzyme_thermodynamics/enzyme_introduction.html

V. Manufacturing Metrology and Standards for the Health Care Enterprise

Brief description of your initiative and its goals:

Apply proven MEL manufacturing technology and expertise to healthcare systems, biomedical devices and equipment, and biomedical data management

Program Manager:
Ram D. Sriram, 301-975-3507, Sriram@nist.gov

Organization:
National Institute of Standards and Technology/Manufacturing Engineering Laboratory (NIST/MEL)

Customer Need & Intended Impact :
Spending on healthcare in the United States was about 13.2% of the GDP in 2000, which is $1.3 trillion, and continues to grow at the rate of 7.3% per year. This amount will reach $2.8 trillion dollars by 2011 (around 17% of the GDP). These costs are also a major concern for the U.S. industry, as escalating healthcare costs are impeding our ability to compete globally. According to a USA Today article, General Motors spent $4.5 billion on healthcare in 2002, an increase of nearly 9% from 2001. GM sold 8.4 million vehicles in 2002. In effect, healthcare expenses constituted $535 of the price tag of each GM car. This is reiterated in a recent U.S. Department of Commerce report entitled "Manufacturing in America: A Comprehensive Strategy to Address the Challenges to U.S. Manufacturers." This report cites that rising healthcare costs may prove to be detrimental to our manufacturing industry, with testimonials from various industries.

Healthcare and manufacturing share many similar organizational, technological and informational issues. Thus, the healthcare industry as a whole is a customer for the metrology, standard-setting support and technology approaches and solutions that MEL has developed for the manufacturing sector that are transferable or adaptable to the healthcare sector.

The benefit to the healthcare industry will be an infrastructure for the accelerated and enriched development of improved organizational, technological and informational support methodologies for all aspects of health care delivery. NIST's contributions will enable much more effective development and application of biological and medical knowledge to practical problems.

Objectives:
There are two dimensions to the program: (1) Healthcare informatics; and (2) Medical devices. Healthcare informatics deals with all the processes or "software" of the healthcare enterprise: modeling and simulation, design and production, biosurveillance, manufacturing and its associated supply chains, and information and data management both in clinical practice and biological research. Medical devices deal with all the products or "hardware" of the enterprise: the characterization, design, manufacture, testing, and metrology of medical devices at scales ranging from large equipment to nano-scale drug delivery mechanisms.

This program deals with the following objectives following the two dimensions discussed above:

(1) Healthcare informatics.

Enterprise modeling and simulation. Explore the applicability of the modeling and simulation technologies developed in MEL to healthcare systems; explore means for disseminating this information to the shareholders in the healthcare industry.
Design and production of pharmaceuticals. Develop representations of pharmaceutical processes, quality measurement methods, and test equipment standards necessary for the specification, characterization, and data interchange involved in the clinical trials, certification, production testing, and manufacture of pharmaceutical products.
Biosurveillance. Develop information models and integration technologies, classifications of healthcare terminologies and ontologies, interchange specifications and test methods necessary for the acquisition, characterization, standardization, and validation of public health surveillance information and the dissemination and integration of relevant treatment guidelines to improve the detection and response to disease outbreaks and insidious bioterrorism attacks.

Manufacturing and value chain management. Develop interfacing specifications and interaction protocols for integrating manufacturing and e-commerce software solutions into biomedical device value chain and enhance existing standards for device integration.

Clinical informatics. Extrapolate from MEL's experience in information modeling and research supporting information interchange standards development for the manufacturing industry to provide experience, assistance and leadership for related activities in the health care informatics field.
Bioinformatics. Adapt and extend NIST's expertise in information modeling, information interchange and standards development in the manufacturing arena to the field of bioinformatics, leading to synergisms with bioinformatics research and practice and consolidation of the bioinformatics knowledge base.

(2) Medical devices.

Mobility devices. Develop test methods and performance metrics, sensor data, standards and specifications necessary for intelligent assistive devices for wheelchair dependents and the blind.

Hearing devices. Develop test and measurement methods, data, standards and specifications necessary for the characterization, manufacturing, and testing of hearing devices and related diagnostic equipment.
Intelligent assistive surgical devices (medical robots). Work with an American Standards for Testing of Materials (ASTM) committee, the Food and Drug Administration (FDA), medical robotic research groups and University Hospitals for

the establishment of Intelligent Assistive Surgical Devices (Medical Robots) standards. This work will extrapolate on our previous work on industrial robot performance and safety standards, related metrology, instrumentation and artifact and marker design.
Surface characterization of biomedical devices. Develop test procedures for characterizing the surfaces of medical devices that relate to device function and failure behavior.

Meso-micro-biodevices. Assist in the establishment of meso-micro-biodevices standards. Meso scale devices have components with feature sizes of a few mm. Micro scale devices have components with features, which range between 1 mm and 1 micrometer and nano scale devices have components with features, which range between 1 micrometer and 1 nanometer.
Nano-biodevices. Develop protocols for high-resolution imaging of individual components and associated complexes of the constituents of nanoparticle drug delivery systems (NDS). Demonstrate imaging of the cell transfection process with fixed and live cells using such systems. Nano scale devices have components with features which range between 1 micrometer and 1 nanometer.
One page descriptions of all projects (both informatics and devices) within the program and provided in the Appendix. The followings papers/reports on healthcare informatics can be downloaded from http://www.mel.nist.gov/ (click on the publications tab).

1. Ravichandran, V., Lubell, J., Vasquez, G.B., Lemkin, P., Sriram, R.D., Gilliland, G.L., *Ongoing Development of Two-Dimensional Polyacrylamide Gel Electrophoresis Data Standards*, Electrophoresis, 25, 297-308, 2004.
2. Ravichandran, V., Sriram, R.D., Gilliland, G.L., *MitoMorphy: An Alignment and Annotation Tool for Human Mitochondrial DNA Polymorphisms*, Mitochondrian, Vol. 4, 309-312, 2004.
3. Ravichandran, V., Vasquez, G.B., Srivastava, S., Verma, M., Petricoln, E., Lubell, J., Sriram, R.D., Barker, P.E., and Gilliland, G., *Data Standards for Proteomics: Mitochondrial Two-Dimensional Polyacrylamide Gel Electrophoresis Data as a Model System*, Mitochondrian, Volume 3, 327-336, 2004.
4. Ravichandran, V. and Sriram, R.D., Toward Data Standards for Proteomics, Nature Biotechnology, Volume 23, Number 3, pp. 373-376, March 2005.
5. Allen, R. H. and Sriram, D., *The Role of Standards in Innovation*,Special Issue on "Innovation: The Key to Progress in Technology and Society," Journal Technological Forecasting and Social Change, 2000.
6. Szczypinski, P., Sriram, R.D. , Sriram, P., Reddy, D., *Model of Deformable Rings for Aiding the Wireless Capsule Endoscopy Video Interpretation and Reporting*, Proceedings of the International Conference on Computer Vision and Graphics 2004, Warsaw, Poland, September 22-24, 2004.
7. Bock, C, Carnahan, L., Fenves, Steven, Gruninger, Michael, Kashyap, V., Lide, B., Nell, James, Raman, R., Sriram, R.D. , Healthcare Strategic Focus Area: Clinical Informatics, NISTIR 7263, (2005)